Nonparametric estimates of gene x environment interaction using local structural equation modeling

September 4, 2014

Nonparametric Estimates of Gene × Environment Interaction Using Local Structural Equation Modeling

Daniel A. Briley[1], K. Paige Harden[1], Timothy C. Bates[2], Elliot M. Tucker-Drob[1]

[1] Department of Psychology and Population Research Center, University of Texas at Austin

[2] Department of Psychology, University of Edinburgh

Running head: LOCAL ESTIMATION OF G×E

Submit correspondence to: Daniel A. Briley, Department of Psychology, University of Texas at Austin, 108 E. Dean Keeton Stop A8000, Austin, TX 78712-1043. Email: daniel.briley@utexas.edu. Phone: 512-471-1406.

Abstract

Gene × Environment (G×E) interaction studies test the hypothesis that the strength of genetic influence varies across environmental contexts. Existing latent variable methods for estimating G×E interaction in twin and family data require the specification of parametric (typically linear) functions for the interaction effect. The chosen function may, therefore, obscure the underlying shape of the interaction effect, and at times, fail to detect a significant interaction if an improper functional form is selected. In this article, we introduce a novel approach to the behavior genetic toolkit, local structural equation modeling (LOSEM). LOSEM is a highly flexible approach for nonparametrically estimating genetic effects across a continuous range of a measured family-level moderator. This approach opens up the ability to detect and visualize new forms of G×E interaction. We illustrate the approach by using LOSEM to estimate gene × socioeconomic status (SES) interactions for six cognitive phenotypes and compare results to those obtained from a conventional parametric approach. For several phenotypes, LOSEM indicates that shifts in genetic variance occur predominantly in the middle range of SES, rather than continuously and monotonically across SES as has been assumed in conventional parametric approaches. For one phenotype, LOSEM reveals a significant nonlinear gene × SES interaction. We offer recommendations for judicious application of LOSEM and provide scripts for implementing biometric LOSEM models in M*plus* and in OpenMx under R.

Keywords: LOSEM; LOESS; kernel regression; gene × environment interaction, cognitive ability

Nonparametric Estimates of Gene × Environment Interaction Using Local Structural Equation Modeling

Gene × Environment (G×E) interaction studies test the hypothesis that the strength of genetic influence varies across environmental contexts, or equivalently, that environmental effects vary as a function of genotype (Plomin et al., 1977). Twin and family behavior genetic approaches test for G×E by estimating latent biometric variance components, typically additive genetic effects (A), shared environmental effects (C), and nonshared environmental effects (E), and examining whether the magnitudes of these variance components differ at different levels of a measured environmental variable.[1] When the measured environment is composed of a small set of discrete categories, testing for G×E is straightforward. However, in many cases the measured environment is a continuous variable. Existing methods for estimating G×E with continuously measured environmental variables require *a priori* specification of the interaction's functional form (Purcell, 2002). If the wrong function has been specified, inferences may be biased and, at times, G×E effects present in the data may not be detected[TB1][D2], effectively lowering power[DB3][etd4].

In the current paper, we present a nonparametric method for the estimating the shape of G×E interaction in twin and family data and provide scripts for implementing this technique in M*plus* (Muthén & Muthén, 1998-2010) and OpenMx (Boker et al., 2011). This method can help researchers better understand patterns in their data and can improve model selection and testing in the analysis of G×E interaction. In the following sections, we first present extant approaches to estimating G×E interaction in biometric twin and family models when the environmental moderator is measured at the family-level (i.e., is shared by members of the twin pair). We then

[1] In this paper, we focus on measured, family-level moderators that are, by definition, the same across family members. This level of measurement is currently required for the statistical approach we introduce, and we return to this limitation in the discussion.

present the novel approach, illustrate it with a real data analysis application, and finally discuss its strengths and limitations.

The Categorical G×E Model

When the environmental moderator is categorical (e.g. impoverished vs. not impoverished), estimating G×E is a straightforward application of multiple-group structural equation modeling (Neale & Maes, 2005, Chapter 9). In the case of a dichotomous moderator and an ACE model fit to data from monozygotic twins reared together (MZ) and dizygotic twins reared together (DZ), instead of the usual two-group model (one group for MZ twins and a second for DZ twins), a four-group model is fit (with additional groups for "low risk" and "high risk" environments each for MZ and DZ twins). Such a model is represented in Figure 1A. Each of the A, C, and E component paths has two labels (e.g. a_l and a_h) to indicate that the parameter is estimated separately for the low ("l") and high ("h") risk levels of the moderator. To test for G×E, parameters for the low and high risk models are constrained to be equal and compared by a χ^2 test to one in which they are allowed to differ between the environmental exposure groups. If the "a" (or c or e) parameters cannot be constrained to be equal across environmental exposure groups without significant loss of model fit, then G×E is supported, as the genetic or environmental variance estimate (e.g., a^2) significantly differs across groups.

In cases in which the environmental moderator has been measured continuously, a researcher seeking to employ this method must categorize the environmental moderator variable by collapsing ranges of the environment into discrete bins. If there is reason to be specifically interested in discrete levels of environmental exposure, or if a researcher has a strong *a priori* reason to expect a discontinuous G×E effect at a known cut point, this categorical approach may be optimal. Without strong guidance from theory or past research, however, researchers must

make arbitrary or intuitive decisions regarding the number of bins to use and the ranges of the environment to cluster (i.e. the location of the cut points). Such decisions offer experimenter degrees of freedom (Simmons et al., 2011), and may fail to capture important aspects of the interaction.

The Parametric G×E Model

Purcell (2002) introduced an extension of the classical twin model for the analysis of G×E interaction with continuously measured environmental moderators. As depicted in Figure 1B, this parametric G×E model controls for the main effect of the observed moderator on the phenotype (the "moderated mean"). Moreover, it specifies that the regression paths from latent biometric factors (A, C, and E) to the phenotype are parametric functions of the observed moderator. This approach is achieved by treating the moderator as a *definition variable*. This, of course, requires raw data, but allows for row-wise or individual-level likelihood model estimation such that the model-implied means and covariances are estimated for each individual pair. When the regression paths are specified to be linear functions of the moderator (as is depicted in Figure 1B), this specification produces ACE variance component estimates that are quadratic functions of the moderator (as the regression path must be squared in order to produce a variance expectation). When the biometric interaction model is expanded to include both linear and quadratic interactions on the paths (such that the ACE variance estimates are quartic with respect to the moderator), one can test whether genetic variance is an inverted U-shaped curve, with the highest genetic variance in the "average" environment (e.g., Burt et al., 2006). Others (e.g. Turkheimer & Horn, 2014) have endorsed exponential functions.[2] Still others have

[2] "We prefer an exponential function rather than a quadratic one as a model of the variances. Exponential models share with quadratic models the desirable property of being positive, but have the additional advantage of being monotonic uniformly increasing or decreasing with respect to the moderator. Quadratic models of variances are by

considered how to test for G×E when the moderator is not necessarily shared by members of a twin pair, but may differ between twins, thus allowing for the simultaneous consideration of gene-environment correlation (e.g., Johnson, 2007; Medland et al., 2009; Molenaar & Dolan, 2014; Price & Jaffee, 2008; Rathouz et al., 2008; Schwabe & van den Berg, 2014; van der Sluis et al., 2012; van Hulle et al., 2013). We do not recapitulate these theoretical and technical issues here, but simply refer the reader to this previous literature, and note here that these multivariate extensions also model the paths from the biometric components to the phenotypes as parametric functions of the moderator.

LOSEM: LOESS with Latent Variables

As noted above, categorical G×E is of limited utility when an environmental moderator of interest is truly continuous (e.g., socioeconomic status), because this approach either lumps together potentially distinct environmental contexts, risks cutting the data at suboptimal points, or loses information. Parametric G×E solves these problems by retaining the full continuous range of the environmental variable. Yet parametric G×E models can still be limiting in that the functional form (or competing functional forms) of the interaction must be chosen *a priori.* At times, researchers may not have strong theoretical predictions regarding how potential moderating effects play out in particular parts of the environmental range, or they may suspect that the polynomial function they are estimating is not capturing theoretically relevant effects. One intuitive approach might be to categorize a continuous moderator into several bins and apply a categorical G×E approach to get a sense of the shape of the interaction and which bins might be equated without cost. However, if the bins chosen are too broad, they combine over levels of a moderator for which genetic variance may differ substantially. Bins that are too narrow rely on

---

definition parabolic with respect to the moderator, and once again, biometric interaction models are difficult enough to explain without having to account for why a biometric variance first increases, and then decreases, as a function of SES" (Turkheimer & Horn, 2014, p. 44).

small subsets of the data and are likely too noisy to detect significant effects. Moreover, the optimal cut points in the continuous distribution of the moderator by which to best form the discrete bins are unlikely to be known, both due to uncertainty about the shape of the function, and its correct location along the moderator axis which is often measured on scales lacking an absolute reference point. Trying many different binning strategies to see which one "works" inflates the risk of false discovery (Benjamini & Hochberg, 1995).

Local structural equation modeling (LOSEM) is a method developed by Hildebrandt et al. (2009) to nonparametrically estimate differences in structural equation model parameters across different levels of a measured putative moderator. LOSEM is the latent variable analogue of LOESS (LOcal regrESSion), or locally weighted regression (Cleveland & Devlin, 1988), a nonparametric regression method that fits a "smoothed" line (a loess curve) through the cloud of data points. Both methods draw on kernel regression techniques, in which statistical models are locally estimated for kernels of the data (Li & Racine, 2007). In this context, the term kernel refers to a weighting function used to select datapoints to be used in local analyses. In the following sections, we explain how LOSEM can be applied to produce a nonparametric "smoothed" estimate of how genetic and environmental variances differ across the observed range of a measured family-level environment. The use of the LOSEM approach has the potential to illuminate patterns of G×E that may otherwise be obscured, and may help guide researchers toward selecting the most appropriate parametric G×E models.

Step 1: Specify a general model. First, a general biometric structural equation model is specified exactly as would be done in a non-G×E context. Note that although the hypothesis being examined predicts that some of the parameters of this model differ as a function of a moderator variable, this moderator is not included in the general structural equation model.

Typically, in the simplest nonparametric G×E case in which one is interested in whether the paths from the biometric components to a phenotype differ as a function of the moderator, the specified model would simply be a classical univariate twin ACE model. Of course alternate univariate forms are possible, such as a dominant genetic model (ADE), or a model without a shared environmental estimate (AE). Because the nonparametric approach does not require any moderation effects to be explicitly specified in the general model, it is also easily applied to more complex multivariate models (e.g., Cholesky decomposition, correlated factors model, simplex, etc.; Neale & Maes, 2005). The primary parameters of interest are the pathways from the latent genetic and environmental factors to the phenotype, which, when squared, represent the variance accounted for by the ACE components.

Step 2: Select a range of target values of the moderator. Second, a moderator and range of *target* values of the moderator are selected. For instance, one might be interested in characterizing the magnitudes of latent genetic and environmental influences on a phenotype across the socioeconomic status (SES) range from 2 SD below the mean SES to 2 SD above the mean SES. Care should be taken to avoid extremely high and low value of the moderator (e.g., +/- 3 SD), as the effective sample size may become small and the estimates imprecise. If SES is on a *z*-scale, the target values of the moderator would be a vector from -2 to +2. To gain sufficient clarity of the trends, the vector could include increments of .1 or even .01. Importantly, this decision is *not* the same as the decision regarding how many bins to use in a categorical G×E model. The LOSEM approach makes full use of the entire dataset for every model, whereas binning separates data into discrete subsets. By using smaller intervals for the target value of the moderator in the LOSEM approach, one simply reduces the distance between estimates (i.e. the resolution of the trend), but the estimates do not change depending on the interval. Note also that

choosing smaller interval sizes does not reduce the effective sample size as the weighting function does not depend on the interval size (see below). The only tradeoff for choosing very small intervals is computation time.

Step 3: Specify a weight function. Third, the model of interest is estimated once at each target value of the moderator in the vector created under Step 2. At each step, the observations – (rows of data in the model) – are *weighted* by their distance from the target value of the moderator. For instance, individuals for whom the moderator = 1 will be weighted most highly when the target value is 1, but weighted much less when the target value is -1. In this way, every row in the data set is informative at all levels of the target, but observations that are closest to the location of the target value of the moderator are privileged (weighted more highly) compared to distant observations. Thus, if one were interested in characterizing genetic and environmental influences across -2 SD SES to +2 SD SES in increments of .01, a total of 401 ACE models would be estimated. Each model would be based on the full dataset, but would give different weight to the data based on the specified target level of the moderator.

We follow Hildebrandt et al. (2009) and Gasser et al. (2004) in recommending that weights be calculated based a kernel function in which the bandwidth (*bw*) depends on the total sample size (*N* pairs of twins) and the variability of the moderator (SD_M):

$$bw = 2 * N(^{-1/5}) * SD_M$$

This bandwidth selection is based on a rule of thumb designed to minimize and balance the amount of bias (i.e., oversmoothing) and variability (i.e., undersmoothing) in the produced estimates (Li & Racine, 2007). As the bandwidth is progressively expanded, the weighting function approximates a uniform distribution across the moderator, and the "local" results actually weight all of the data equally. In this case, the estimates are biased in the sense that they

do not capture the nonparametric trend. As the bandwidth is progressively shrunk, the weighting function only considers data at a specific level of the moderator. When these estimates are combined to produce the nonparametric trend, variability in the estimates is maximized. We return to alternative specifications of the weighting function in the discussion.

The scaled distance (z_i) between the value of *M* for each individual *i* and the target value of *M* is then scaled according to *bw*:

$$z_i = (M_i - \text{target } M) / \text{bw}$$

The kernel weights (K) for each individual *i, for each target value of M,* is then calculated based on this distance, and re-scaled as a final weight (W) to vary between 0 and 1:

$$K = (1 / \sqrt{2\pi}) * \exp(-z_i^2 / 2)$$

$$W = K / .399$$

Figure 2 shows example weighting distributions. The distribution of weights varies as a function of sample size and the standard deviation of the moderator. Larger sample sizes and smaller standard deviations of the moderator both result in weighting distributions more tightly focused around the target moderator value. Figure 2A illustrates weighting distributions based on data used in the current study ($N = 650$, moderator SD = 1).[3] Figure 2B shares the moderator SD of Figure ~~1B~~2B, but is based on a ten times larger sample ($N = 6500$) to demonstrate how the distribution of weights shrinks with larger samples. The *bw* parameter is the primary determinant of the width of the weighting distribution. If desired, this parameter is easily manipulated by the researcher to produce different levels of smoothing.

Step 4: Run the model for each target value of the moderator and compile estimates. Finally, the biometric model of interest is estimated once at each target value of the moderator,

[3] Due to ECLS-B data regulations, all sample sizes are rounded to the nearest 50.

each time weighting the observations by their distance from the current target. To examine the obtained nonparametric G×E curve, the user may then plot parameters of interest (e.g., the squared additive genetic path from the A factor to phenotype) as a function of the value of the target moderator. This approach renders the nonparametric function of the genetic variance moving smoothly across values of the environmental moderator.

The LOSEM approach to G×E is shown as a path diagram in Figure 1C. Each parameter is estimated at each of a range of target values of the continuous moderator (in Figure 1C we specify this range in terms of[etd5] ~~z-score standard deviations~~*z* units above and below ~~from~~ a mean of 0), and this information is aggregated to yield a nonparametric function of the parameter estimates across the chosen range of the moderator. In other words LOSEM involves running a large number of models — one for each "target" value of the moderator —the estimates from all models are combined into a nonparametric representation of how parameters differ across the range of the moderator.

Work flow and implementation in M*plus* and *R*. The online supplement includes example scripts to implement LOSEM. For analysts using M*plus* (Muthén & Muthén, 1998-2010), automating the multiple models that need to be run can be accomplished using the "MplusAutomation" package in *R* (Hallquist, 2011; Ihaka & Gentleman, 1996). This package includes commands to (1) create multiple modified input files based on a template, (2) run all of the input models, and (3) extract and combine model parameters from the output files (see Appendix A and B for sample scripts). *R* can then be used to extract model parameters and bind these into a dataset across target levels of the moderator with associated model parameters and standard errors. This dataset can then easily be used to plot ~~dynamic~~ nonparametric G[etd6]×E interaction trends. Using OpenMx (Boker et al., 2011), the functionality of *R* can be used to

accomplish similar tasks directly (see Appendix C for sample scripts). These packages make it extremely easy to run, extract, and aggregate all of the necessary models and parameter estimates. The whole process can take as little as 15 minutes.

In the sections that follow, we demonstrate the power of this approach by re-analyzing gene × SES findings and show a potentially novel pattern of result that would have been obscured had LOSEM not been employed. Finally, we make several methodological recommendations concerning the judicious application of LOSEM.

Example: Childhood SES moderation of Genetic Effects on Cognitive Ability in the ECLS-B

We have previously used LOSEM in a study of how birth cohorts differ in genetic influences on fertility behavior (Briley et al., *under review*[D7][etd8]) and in a study of how the relation between pubertal timing and depression varies as a function of SES (Mendle et al., *under review*). In both of these cases, we expected nonlinear G×E trends, but it was unclear what the exact functional form was. LOSEM allowed us to explore the data and make informed analytic choices. Here we present another example of LOSEM for the analysis of G×E interaction using data from the Early Childhood Longitudinal Study – Birth Cohort (ECLS-B; Snow et al., 2009). Previous publications have reported results of parametric gene × SES interaction analyses in this dataset. Tucker-Drob et al. (2011) reported that longitudinal increases in mental ability between 10 months and 2 years were more heritable among children being raised in higher SES families. Rhemtulla and Tucker-Drob (2012; also see Tucker-Drob & Harden, 2012) reported that age 4 math, but not age 4 reading, was more heritable among children being raised in higher SES families.

These results are consistent with a bioecological model, in which resource-rich environments allow for personal interests, preferences, desires, and temperaments to play a large role in development (e.g., Bronfenbrenner & Ceci, 1994; Tucker-Drob et al., 2013). However, alternative theoretical models have been proposed in which there is a nonlinear relation between environmental circumstances and the genetic variance of cognition (e.g., Scarr, 1992; Turkheimer & Gottesman, 1991). Under a model of the "average expectable environment" (Scarr, 1992, p. 5), genetic variance is predicted to increase as the environment transitions from bad to average, but then plateau following. According to this perspective, there is a dramatic difference between growing up in poverty and growing up in the middle class, but a less appreciable difference between growing up middle class and wealthy. By applying LOSEM to nonparametrically model the shape of the G×SES interactions, we seek to determine whether SES-related increases in genetic variance occur throughout the range of the SES distribution or are confined to a specific range of the SES distribution. We apply LOSEM to all available cognitive phenotypes: 10 months Bayley mental development, age 2 years Bayley mental development, age 4 years math and reading readiness, and kindergarten math and reading achievement. For methodological details on the ECLS-B sample and measurement of these phenotypes, including sample statistics, please see Rhemtulla and Tucker-Drob (2012), Tucker-Drob and Harden (2012), Tucker-Drob (2012), and Tucker-Drob et al. (2011).

Results

Figure 3 compares LOSEM results with traditional Purcell (2002) results. The first two columns present variance accounted for in the phenotype by ACE factors. Dotted lines represent +/- 1 standard error of the estimate. The last two columns present the main effect of the moderator. For the LOSEM approach, the graph plots the estimated mean of the phenotype (i.e.,

the estimated twin mean of cognitive ability at SES = -2 to +2). For the Purcell (2002) approach, the graph plots the regression parameter for the main effect of SES. Table S1 and Supplementary Files S1-2 present parameter estimates and model fit statistics for models fit for the current study and a more complete analytic description. In the context of the Purcell (2002) model, we found significant genetic interaction terms for age 2 Bayley (a' = .193, $p < .001$) and age 4 math (a' = .164, $p < .001$). Significant interaction terms for the shared environment (c' = .106, $p < .01$) and the nonshared environment (e' = .075, $p < .001$) were found for age 4 reading. No other interaction terms were significant for the standard application of the Purcell (2002) model. Higher levels of SES were associated with higher levels of ability. These coefficients ranged from .030 (*n.s.*) to .476 ($p < .001$, see Table S1).

Across most models, there was generally good consilience between the LOSEM results and the Purcell (2002) specification. For example, the estimate of genetic variance from the two approaches correlated very strongly at -2 SES ($r = .98$), -1 SES ($r = .94$), -.5 SES ($r = .94$), +.5 SES ($r = .96$), and +1 SES ($r = .91$), but less strongly for +2 SES ($r = .55$). This is primarily due to the large nonlinear shifts observed at high levels of SES using LOSEM, particularly for kindergarten reading. More descriptively, the approaches largely agree on the directional trend of the variance components. For age 1 and age 2 Bayley, shared environmental influences decrease, nonshared environmental influences decrease slightly, and genetic influences increase with increasing SES across both analytic approaches. The general directional trends are also very similar for age 4 math and reading. The results are less consistent for kindergarten math and reading as the LOSEM results imply fluctuating levels of genetic and environmental influences, but the Purcell (2002) results imply very little change in parameters across different SES levels.

Despite the general agreement between approaches in broad trends, a few substantial differences are evident. Most notably, the linear Purcell (2002) interaction model for kindergarten reading ability (Figure 3F) is clearly miss-specified. The LOSEM results indicate relatively low genetic variance at low SES, a spike in genetic variance near average levels of SES, and a steep decline in genetic variance at high levels of SES. The linear specification of the Purcell (2002) model indicates that there is essentially no difference in genetic variance across SES, obscuring rather large differences apparent in the LOSEM results. Figure 4 presents a specification of the Purcell (2002) model that includes a quadratic interaction term, which is significant for genetic influences (see Table S1).

As discussed earlier, it may be theoretically relevant where in the research range of interest of the moderator most of the increases or decreases in genetic variance occur. For example, genetic variance may increase by a total of .4 between SES of -2 and +2, but .3 of that increase may occur between SES of -0.5 and +0.5 and only .1 between SES of +1 and +2. This would be more in line with the "average expectable environment" model, which predicts the largest difference to be between bad and good-enough environments, not between good and excellent environments. The LOSEM approach easily captures this important information. On the other hand, the Purcell (2002) model, due to its specification, tends to predict more extreme increases for more extreme values of the moderator. At least for the six phenotypes under investigation in the current study, this does not seem well-warranted.

Table 1 compares differences in the magnitude of genetic variance across meaningful levels of the moderator for each analytic approach. In particular, we were interested in whether interaction effects were concentrated at the low-range (SES from -2 SD to -1 SD, $\Delta\, a^2$ low), mid-range (SES from -0.5 SD to +0.5 SD, $\Delta\, a^2$ mid) or high-range (SES from +1 SD to +2 SD, $\Delta\, a^2$

high).[4] Both approaches indicate that there is very little increase in genetic variance across the low-range of SES. Focusing on the LOSEM approach, genetic variance increases to a greater extent in the mid-range than in the high-range for all phenotypes except age 1 Bayley and age 4 reading (for which there was essentially no interaction). Turning toward the Purcell (2002) results, this trend is not evident as the increase in genetic variance is always higher for the high-range of SES (as required by the model).

Recommendations for employing LOSEM

In this section, we offer initial recommendations on effective ways to use the LOSEM approach to inform studies of G×E interaction.

Define the research range of interest. The research range of interest refers to the span of the moderator that is under investigation (Roisman et al., 2012). For example, the plots in Figure 3 are based on a research range of interest between - 2 SD and + 2 SD of SES, a region that contains nearly all of the empirical observations and does not extend to regions of no data availability. Using a parametric approach, it would be analytically feasible to explore the range from -8 SD to -4 SD of SES, but this range would extrapolate well beyond the data. Similarly, LOSEM trends identified where data density is sufficient could be extrapolated to nonsensical regions. Applying LOSEM to these regions would likely result in nonsensical estimates as the majority of the data would receive relatively equal, and very small, weight in the analysis. These practices should be avoided. Of course, researchers should interpret results based on sufficient data points and should take care to report on their moderator in reference to a general standard (i.e., whether the moderator spans from bad to normal, such as child maltreatment to no child maltreatment, or from bad to good, such as is the case for most standardized measures of SES in representative samples).

[4] Of course, such an approach is inadequate to capture many of the nonlinearities found in the data.

Get the main effects right. Just as the parametric G×E approach requires that the shape of the interaction effect conform to a parametric function, it also requires that the main effect of the moderator conform to a parametric function. The LOSEM approach estimates the main effect of the moderator on the phenotype nonparametrically. When comparing parametric and LOSEM approaches, it is important to first scrutinize the extent to which the main effects across levels of the moderator are consistent across the two methods prior to interpreting the interaction effects. If the main effects differ, the interaction component may also differ, as the biometric components in both models, including the interaction component model phenotypic variance that is *unique* of the moderator. In a situation in which the main effects from the two approaches are not in close agreement, one could use the main effect trend line produced by the LOSEM approach (column 3 of Figure 3 and column H of Supplementary File 2) to residualize the phenotype using this trendline prior to implementing the parametric model. In this case, the main effect of M and the Y intercept could be set to zero ~~the estimated main effect of the moderator~~ in the parametric model. ~~should be very close to zero[TB9], and the mean trends in the phenotype would be identical to the LOSEM trends.~~ This ~~will yield a set of Purcell-type~~would enable for parametric and nonparaemtic modeling of variance in the exact same residuals, such that direct comparisons between the two approaches could be made. ~~results for variance components that may validly be compared to the LOSEM results[TB10]~~[etd11]~~.~~

Choose the right baseline model. Standard approaches to model fitting/trimming (e.g., Neale et al., 1989) can guide the selection of reduced biometric models (e.g., an AE model over an ACE model) or an alternative model (e.g., one including D rather than C variance components). In the case of LOSEM, the possibility exists for reducing or otherwise varying models differently at each target value of the moderator. Such locally-distinct genetic

architectures are unlikely to have biological validity. Different genetic architectures would imply mechanisms that exist in the organism exerting not just varying influence on the phenotype, but are actually absent at some levels of the moderator. For this reason, we recommend that the same variance components are modeled across all levels of the moderator, with differences in magnitude being the primary focus. Still, it is conceivable that there are highly complex processes at play that give rise to a situation, for example, in which dominant genetic influences on a phenotype are only manifest at certain levels of a moderator. Ultimately, this is a data- and topic-specific question, and the near-endless modeling possibilities must be tempered by the principle of parsimony.

Choose the right tool from the toolkit. The major advantage of using a nonparametric exploratory approach, such as LOSEM, is the ability to detect (or to rule out) nonlinear G×E interactions. In the current application of LOSEM, we detected model misspecification for kindergarten reading ability and corrected this by applying a more appropriate interaction function (Figure 4). This pattern would not be detected given typical model fitting approaches, but is very easily noticed when nonparametric approaches are used to inform model selection. Of course, additional data are necessary to evaluate the replicability of the non-linearity, which was only observed for one measure (reading) and at one developmental period (Kindergarten). It is unclear whether other G×E interaction studies may have reported ~~null (or positive)~~biased [TB12][etd13]results simply due to inappropriate statistical models. Incorporating flexible, nonparametric approaches as a data analytic step can help avoid such pitfalls.

Examine differences in the magnitude of variance across meaningful ranges of the moderator and the proportion affected. To supplement basic visual inspection of LOSEM trends, differences in the magnitudes of variance offers a convenient way to quantify how

quickly genetic or environmental influences shift over meaningful levels of the moderator. For example, Table 1 demonstrates how this approach can help guide interpretation of trends. Further, Roisman et al. (2012) suggested calculating the "proportion affected" when evaluating the shape and importance of candidate G×E interaction results. Individuals are "affected" by the interaction if they experience a level of the environment beyond the crossover point of a candidate G×E interaction (i.e., the point of the moderator at which two genotypes appear equivalent on a phenotype). The region beyond this point indicates that genotypes are responding differently to the environment. They argued that if 16% or 2% of the sample falls above this point, then that would provide good or speculative evidence, respectively, for the practical importance of an interaction effect. This convention was suggested based on reference to a normal distribution in which 16% and 2% of the sample would be 1 and 2 SD above the mean, respectively. In the current context, the spike in genetic variance for age 4 math occurs at SES of +1.5 SD, indicating that approximately 7% of the sample is "affected" by the spike. Thus, there is decent evidence that this increase in genetic variance has a meaningful effect (i.e., accounting for approximately half of the total increase in genetic variance, see Table 1) on a meaningful proportion of the sample.

~~Discussion~~[TB14][D15] Conclusions and Future Directions

We have demonstrated the utility of a novel approach to analyzing G×E interaction results. LOSEM produces flexible~~, dynamic~~ nonparametric estimates of trends G×E interaction that can detect nonlinearities and inform subsequent confirmatory model fitting. We applied this approach to a highly studied effect with widely used data to make novel insights concerning trends found in the data. We plotted nonparametric estimates of genetic, shared environmental, and nonshared environmental variance across levels of SES in the ECLS-B sample for six

cognitive ability phenotypes. Using the LOSEM approach, we detected an inverted-U shape curve for the genetic variance of kindergarten reading ability. Following up this approach with a standard parametric model that included a quadratic term (Purcell, 2002), we confirmed that this nonlinearity was statistically significant. As mentioned previously, this result for a single phenotype at a single age requires additional replication and investigation before it can inform theory, but the process of discovery represents a key strength of LOSEM.

Additionally, we used the flexible LOSEM results to probe where in the SES distribution the majority of the differences in magnitude of genetic variance occur. For several phenotypes, the majority of the G×SES interaction occurred in the transition from somewhat bad to somewhat good environments with almost no increase associated with the good to excellent [etd16]range. Again, this trend would be completely missed if relying solely on parametric, linear models. Of course, the current study is primarily concerned with displaying the utility of the novel LOSEM approach for G×E interaction studies. Much more empirical evidence will be needed to evaluate the exact functional form of this interaction across different ages and cognitive phenotypes.

As with all exploratory approaches, LOSEM has potential pitfalls. Exploratory data analysis opens up researcher degrees of freedom that might allow for inappropriate manipulation of data to capitalize on noise (Simmons et al., 2011). For example, LOSEM results could be used to find just the "right" points of the moderator to dichotomize or categorize different groups. A related pitfall would be to over-interpret minor deviations of the LOSEM trends as meaningful effects. We have provided some recommendations for avoiding this pitfall, such as using the proportion affected by the trend and following LOSEM analyses with confirmatory approaches.

Interpretation of LOSEM must balance the detection of meaningful nuance from random noise. This balance is primarily determined by the *bw* parameter. When this parameter is increased, noise in the estimates is reduced, but ~~this increases the bias of the estimates~~more nuanced micro-trends may be missed. When shrunk, ~~noise in the estimates is increased, but~~ the estimates conform closely to local subsets of the data, thus increasing the capability to pick up on nuanced trends, but also increasing the chance of picking up on statistical noise. ~~and have greater variance. This bias-variance~~This tradeoff is inherent in kernel regression methodology (Li & Racine, 2007). We have followed the recommendation of Hildebrandt et al. (2009) and Gasser et al. (2004) in calculating *bw* based on the sample size and standard deviation of the moderator. When certain assumptions [etd17]hold perfectly, it can be shown that this produces the ideal bandwidth (Li & Racine, 2007). As discussed extensively in previously published work on nonparaemteric regression methods, a number of other data driven methods exist for choosing the optimal *bw* (Bowman, 1984; Rudemo, 1982; ~~. These include "plug-in" methods in which several different *bw*'s are used to determine if fit differs[etd18], cross-validation in which a *bw* that minimizes the integrated mean square error is found [etd19](Bowman, 1984; Rudemo, 1982~~Hurvich et al., 1998~~), and minimization of a corrected Akaike's Information Criterion designed specifically for nonparametric regression (Hurvich et al., 1998[etd20])~~), along with adaptive bandwith approaches, ni which local estimates are weighted by a constant number of nearby datapoints. Each of these methods may provide slightly different values for *bw* and therefore possibly produce substatntively different ~~end results~~trend estimates. Additionally, we followed Hildebrandt et al. (2009) and Gasser et al. (2004) in recommending the kernel function follow a Gaussian distribution, but a number of other functional forms are available.

~~These include uniform, triangular, quartic, bisquare, or any other desirable functional form. Alternatively, an adaptive bandwidth could be used in which a specific number of nearby data points are included in every analysis. This has the desirable feature of ensuring that every local estimate is based on sufficient data and is less susceptible to "edge" effects (i.e., suspect estimates at extreme ends of the moderator which are not symmetrically weighted simply due to the distribution of the moderator).~~

In its current form, LOSEM is primarily an exploratory analytic tool, but inferential applications would be desirable. Hülür et al. (2011) used permutation as an inferential tool for LOSEM in a non-behavioral genetic context. They created 1,000 permuted datasets in which they randomly assigned each datapoint a value of the moderator and estimated the structural parameter of interest in each dataset. They compared the trend derived from the observed data to the 95% confidence interval of the trend from the permuted data across the moderator. They interpreted estimates derived from the observed data that did not overlap with the 95% confidence interval of the permuted data as evidence of statistical significance. This approach is limited in the sense that it does not produce a single test statistic. Racine (1997) provides a ~~robust and consistent~~ significance test for nonparametric regression which is also based on resampling techniques. This approach generates a test statistic of whether the partial derivative of a given nonparametric regression coefficient is zero (i.e., the null hypothesis) or is greater than zero (i.e., the alternative hypothesis). [etd21]Alternatively, one could test whether the LOSEM results significantly differ from a parametric G×E model if the null distribution is based on the parametric results rather than a model of no interaction. Because the distribution of the test statistic is unknown, pivoted bootstrap resampling is used to generate the null distribution, and a single significance test can be calculated. [etd22]~~Both of these~~Resamping approaches ~~are~~ would be

computationally intensive, as they. T hey would require performing LOSEM on the observed data, generating many resamplings of the data, and then performing LOSEM on each resampled dataset. Further, Racine's (1997) significance test does not directly translate to a structural equation modeling context, but similar principles could be applied. It would be desirable if a test could be constructed based on LOSEM results (e.g., Figure 3) that did not require the use of raw data[D23].

A major limitation of the LOSEM approach is that it requires the environmental moderator to be measured at the family-level. Quantitative behavior genetic methods use the sibling pair as the unit of analysis, and the weighting function must be applied at this level. Therefore, the LOSEM approach, in its current form, is unable to estimate G×E for moderators that vary within families. Several papers have developed and scrutinized parametric G×E methods for moderators that vary within families (Rathouz et al., 2008; van Hulle et al., 2013; van der Sluis et al., 2012), which allow for modeling of gene-environment correlation. Future efforts to develop LOSEM methods to handle such data structures would be highly valuable.

In conclusion, Conclusion

LOSEM can be a valuable tool in the behavior genetic toolkit for probing G×E interactions. As researchers have successfully adopted LOESS approaches to regression to explore and visualize data, LOSEM can be applied to behavior genetic data to detect nonlinearities or discontinuities of trends that would otherwise be missed. In the online supplement, we provide scripts for implementing LOSEM in M*plus* and in OpenMx. We encourage researchers to apply LOSEM to better understand the complex interplay between genetic and environmental influences.

Acknowledgements

This research was supported by National Institutes of Health (NIH) research grant R21-HD069772 to Dr. Tucker-Drob and R21-AA020588 to Dr. Harden. Daniel A. Briley was supported by NIH training grant T32HD007081. The Population Research Center at the University of Texas at Austin is supported by NIH center grant R24HD042849. [Add Bates Acknowledgements]

Conflict of Interests

All authors declare that they have no conflict of interest.

References

Benjamin Y, Hochberg Y (1995) Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J R Stat Soc Series B Stat Methodol* 57(1): 289-300.

Boker S, Neale M, Maes H, Wilde M, Spiegel M, Brick T, … Fox J (2011) OpenMx: An open source extended structural equation modeling framework. *Psychometrika,* 76: 306-317.

Bowman AW (1984) An alternative method of cross-validation for the smoothing of density estimates. *Biometrika* 71:353-360.

Briley DA, Harden KP, Tucker-Drob EM (*under revi~~sion~~ew*). Genotype × cohort interaction on completed fertility and age at first birth. ~~*Behavior Genetics.*~~

Bronfenbrenner U, Ceci SJ (1994) Nature-nurture reconceptualized in developmental perspective: A bioecological model. *Psychol Rev* 101:568-586.

Burt SA, McGue M, DeMarte JA, Krueger RF, Iacono WG (2006). Timing of menarche and the origins of conduct disorder. *Arch Gen Psychiatry* 63(8):890-896.

Cleveland WS, Devlin SJ (1988) Locally-weighted regression: An approach to regression analysis by local fitting. *J Am Stat Assoc* 83(403):596-610.

Gasser T, Gervini D, Molinari L (2004) Kernel estimation, shape-invariant modeling and structural analysis. In R Hauspie, N Cameron, L Molinari (Eds.) *Methods in human growth research*. Cambridge University Press, pp 15-33.

Hallquist M (2011) MplusAutomation: Automating Mplus model estimation and interpretation. *R package version 0.5*.

Hildebrandt A, Wilhelm O, Robitzsch A (2009) Complementary and competing factor analytic approaches for the investigation of measurement invariance. *Rev Psychol* 16(2):87-102.

Hülür G, Wilhelm O, Robitzsch A (2011). Intelligence differentiation in early childhood. *J Individ Differ* 32(3):170-179.

Huvich CM, Simonoff JS, Tsai C-L (1998). Smoothing parameter selection in nonparametric regression using an improved Akaike information criterion. *J R Stat Soc Series B Methodol* 60:271-293.

Ihaka R, Gentleman R (1996) R: A language for data analysis and graphics. *J Comput Graph Stat* 5:299-314.

Johnson W (2007) Genetic and environmental influences on behavior: capturing all the interplay. *Psychol Rev* 114(2):423-440.

Li Q, Racine J (2007) *Nonparametric econometrics: Theory and practice*. Princeton University Press.

Medland SE, Neale MC, Eaves LJ, Neale BM (2009) A note on the parameterization of Purcell's G× E model for ordinal and binary data. *Behav Genet* 39(2):220-229.

Mendle J, Moore SR, Harden KP (*under revision*) Puberty, socioeconomic status, and depressive symptoms in adolescent girls: Evidence for genotype × environment interactions. *Clin Psychol Sci*.

Molenaar D, Dolan CV (2014) Testing systematic genotype by environment interactions using item level data. *Behav Genet*. Advance online publication.

Muthén LK, Muthén BO (1998-2010) *Mplus user's guide* (6th ed.). Los Angeles, CA: Muthén and Muthén.

Neale MC, Heath AC, Hewitt JK, Eaves LJ, Fulker DW (1989) Fitting genetic models with LISREL: Hypothesis testing. *Behav Genet* 19(1):37-49.

Neale MC, Maes HHM (2005). *Methodology for genetic studies of twins and families*. Dordrecht, The Netherlands: Kluwer Academic Publishers.

Plomin R, DeFries DC, Loehlin JC (1977). Genotype-environment interaction and correlation in the analysis of human behavior. *Psychol Bull* 84(2):309-322.

Price TS, Jaffee SR (2008) Effects of the family environment: gene-environment interaction and passive gene-environment correlation. *Dev Psychol* 44(2):305-315.

Purcell S (2002) Variance components models for gene-environment interaction in twin analysis. *Twin Res* 5:554-571.

Racine J (1997) Consistent significance testing for nonparametric regression. *J Bus Econ Stat* 15(3):369-378.

Rathouz PJ, Van Hulle CA, Rodgers JL, Waldman ID, Lahey BB (2008) Specification, testing, and interpretation of gene-by-measured-environment interaction models in the presence of gene–environment correlation. *Behav Genet* 38(3):301-315.

Rhemtulla M, Tucker-Drob EM (2012) Gene-by-socioeconomic status interaction on school readiness. *Behav Genet* 42:549-558.

Roisman GI, Newman DA, Fraley RC, Haltigan JD, Groh AM, & Haydon KC (2012) Distinguishing differential susceptibility from diathesis-stress: Recommendations for evaluating interaction effects. *Dev Psychopathol* 24:389-409.

Rudemo M (1982) Empirical choice of histograms and kernel density estimates. *Scand Stat Theory Appl* 9:65-78.

Scarr S (1992) Developmental theories for the 1990s: Development and individual differences. *Child Dev* 63:1-19.

Schwabe I, van den Berg SM (2014) Assessing genotype by environment interaction in case of heterogeneous measurement error. *Behav Genet*. Advance online publication.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychol Sci* 22:1359-1366.

Snow K, Derecho A, Wheeless S, Lennon J, Rosen J, Rogers J, . . . Einaudi P (2009) *Early Childhood Longitudinal Study, Birth Cohort (ECLS-B), kindergarten 2006 and 2007 data file user's manual (2010-010)*. Washington, DC: U.S. Department of Education, National Center for Education Statistics, Institute of Education Sciences.

Tucker-Drob EM (2012) Preschools reduce early academic achievement gaps: A longitudinal twin approach. *Psychol Sci* 23:310-319.

Tucker-Drob EM, Briley DA, Harden KP (2013) Genetic and environmental influences on cognition across development and context. *Curr Dir Psychol Sci* 22:349-355.

Tucker-Drob EM, Harden KP (2012) Learning motivation mediates gene-by-socioeconomic status interaction on early mathematics achievement. *Learn Individ Differ* 22:37-45.

Tucker-Drob EM, Rhemtulla M, Harden KP, Turkheimer E, Fask D (2011) Emergence of gene-by-socioeconomic status interaction on infant mental ability between 10 months and 2 years. *Psychol Sci* 22:125-133.

Turkheimer E, Gottesman II (1991) Individual differences and the canalization of human behavior. *Dev Psychol* 27:18-22.

Turkheimer E, Horn EE (2014). Interactions between socioeconomic status and components of variation in cognitive ability. In D Finkel and CA Reynolds (Eds.) *Behavior genetics of cognition across the lifespan* (pp.41-68). New York: Springer.

van der Sluis S, Posthuma D, Dolan CV (2012) A note on false positives and power in G×E modelling of twin data. *Behav Genet* 42(1):170-186.

van Hulle CA, Lahey BB, Rathouz PJ (2013) Operating characteristics of alternative statistical methods for detecting gene-by-measured environment interaction in the presence of gene–environment correlation in twin and sibling studies. *Behav Genet* 43(1):71-84.

Figure Captions

Figure 1. Path diagrams representing each type of G×E model. In all models, latent additive genetic (A), shared environmental (C), and nonshared environmental (E) factors with are estimated for a phenotype for twin1 (Y1) and a phenotype for twin2 (Y2). The A factors correlate at 1.0 for monozygotic twins and at 0.5 for dizygotic twins. The C factors correlate at 1.0, and the E factors are uncorrelated. **A.** Categorical G×E in which separate parameters are estimated for low risk (a_l, c_l, e_l, and ~~υ~~μ_l) and high risk (a_h, c_h, e_h, and ~~υ~~μ_h) environments. **B.** Parametric G×E model in which the focal pathways are specified to be a linear combination of parameters representing main effects (a, c, e) and interaction terms (a', c', and e') of the ACE components with the moderator (M). ~~The phenotypes are regressed on M~~The main effect of M is represented as a "moderated mean" (~~B_1~~b_1). The intercept of the phenotype is also estimated ~~, and residual means are estimated~~ (~~B_0~~b_0). **C.** Nonparametric LOSEM G×E model in which local parameters for each level of the moderator are estimated ($\hat{a}_M$, $\hat{c}_M$, $\hat{e}_M$, $\hat{\mu}_M$~~$\hat{u}_{[etd24]}$M~~), noting the circumflex refers to the fact that these parameters are based on weighted data rather than data precisely at the level of M. The subscript [-z…0…+z] refers to the fact that the parameters are actually vectors that include weighted estimates from a lower bound of M to an upper bound of M. Here, we assume that standardized values of M are used and denoted as z.[etd25]

Figure 2. Example distributions of weighting variable (y-axis) at three target levels of the moderator (x-axis). Data closer to the target level of the moderator carries more weight in the analysis. The distribution around the target is smaller with larger sample size and smaller standard deviation of the moderator. **A.** Distribution for the current analysis based on data from ECLS-B (N = 650, SD = 1). **B.** Distribution for hypothetical analysis based on data from ECLS-B with ten times the number of participants.

Figure 3. Comparison of LOSEM and Purcell (2002) gene × socioeconomic status results for cognitive ability measures from ECLS-B. **A.** Age 10 months Bayley. **B.** Age 2 years Bayley. **C.** Age 4 years Math Achievement. **D.** Age 4 years Reading Achievement. **E.** Kindergarten Math Achievement. **F.** Kindergarten Reading Achievement.

Figure 4. LOSEM, linear Purcell (2002), and nonlinear Purcell (2002) model for Kindergarten reading achievement.

Table 1. Comparison of differences in the magnitude of genetic variance across levels of SES between LOSEM and Purcell (2002)

	LOSEM				Purcell			
Phenotype	Δ a^2	Δ a^2 low	Δ a^2 mid	Δ a^2 high	Δ a^2	Δ a^2 low	Δ a^2 mid	Δ a^2 high
Age 10 months Bayley	.162	-.025	-.032	.246	.027	-.005	.007	.019
Age 2 years Bayley	.322	-.044	.372	-.133	.612	.041	.153	.265
Age 4 years Math	.412	-.003	.189	.186	.576	.063	.144	.225
Age 4 years Read	.112	.046	-.102	.092	.059	.013	.014	.017
K Math	-.025	.022	.096	-.148	-.055	-.014	-.014	-.014
K Read	-.204	.028	.212	-.387	.051	.012	.013	.014

Notes. K = kindergarten. Δ a^2 = (a^2 at SES +2) – (a^2 at SES -2). Δ a^2 low = (a^2 at SES -1) – (a^2 at SES -2). Δ a^2 mid = (a^2 at SES +0.5) – (a^2 at SES -0.5). Δ a^2 high = (a^2 at SES +2) – (a^2 at SES +1). Linear specification of Purcell (2002) used for all comparisons.

Figure 1

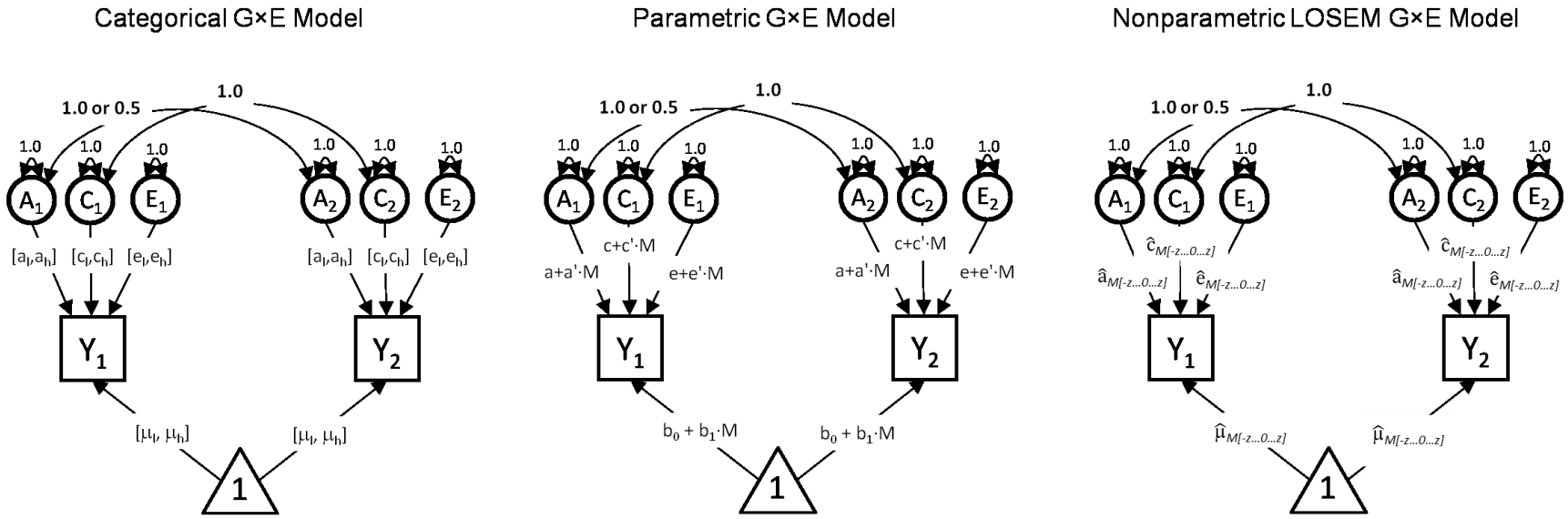

Categorical G×E Model
Parametric G×E Model
Nonparametric LOSEM G×E Model
1.0
1.0 or 0.5
A1
C1
E1
A2
C2
E2
[al,ah] [cl,ch] [el,eh]
Y1
Y2
[μl, μh]
1
c+c'·M
a+a'·M
e+e'·M
b0 + b1·M
M[-z...0...z]

Figure 2

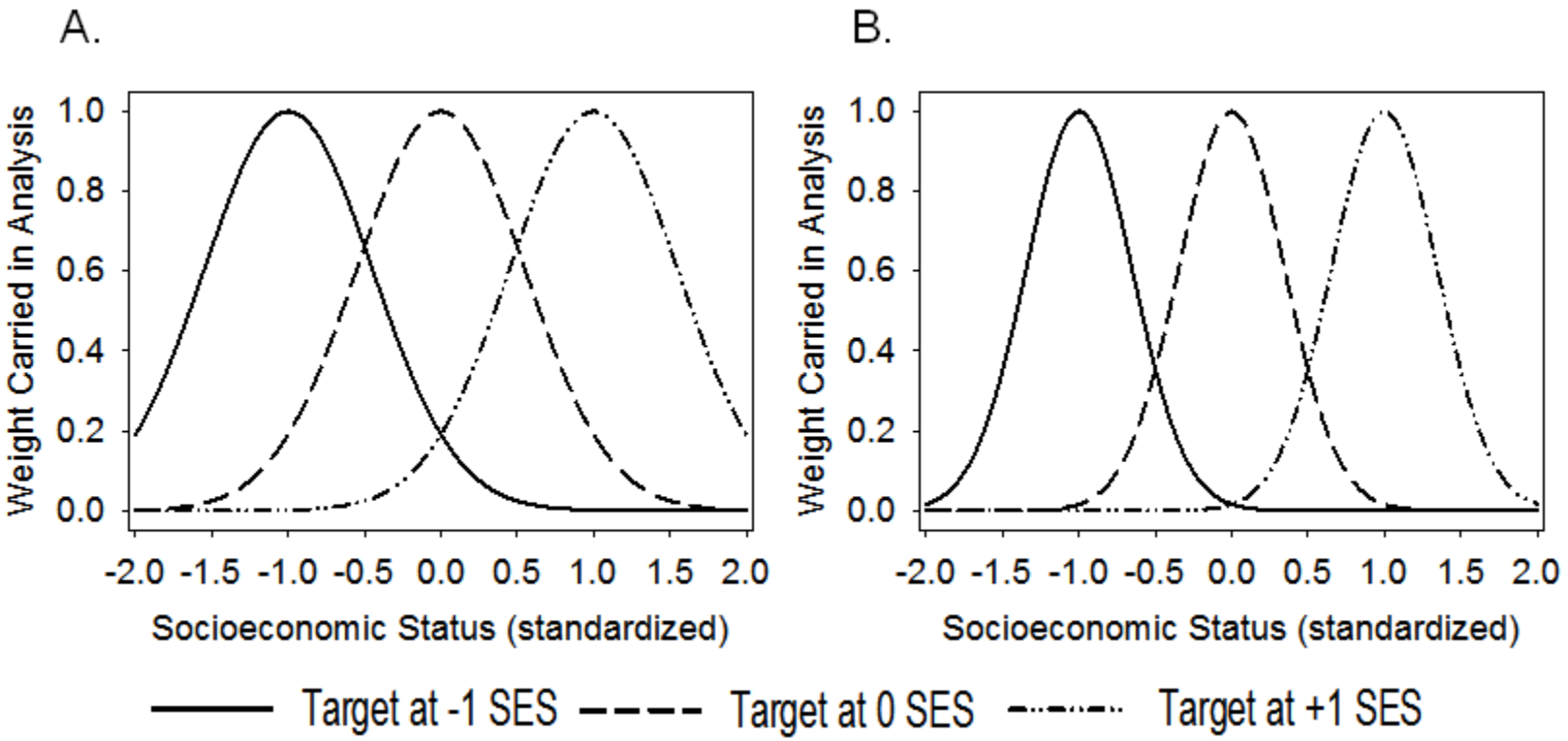

A.
B.
Weight Carried in Analysis
1.0
0.8
0.6
0.4
0.2
0.0
-2.0 -1.5 -1.0 -0.5 0.0 0.5 1.0 1.5 2.0
Socioeconomic Status (standardized)
Target at -1 SES
Target at 0 SES
Target at +1 SES

Figure 3

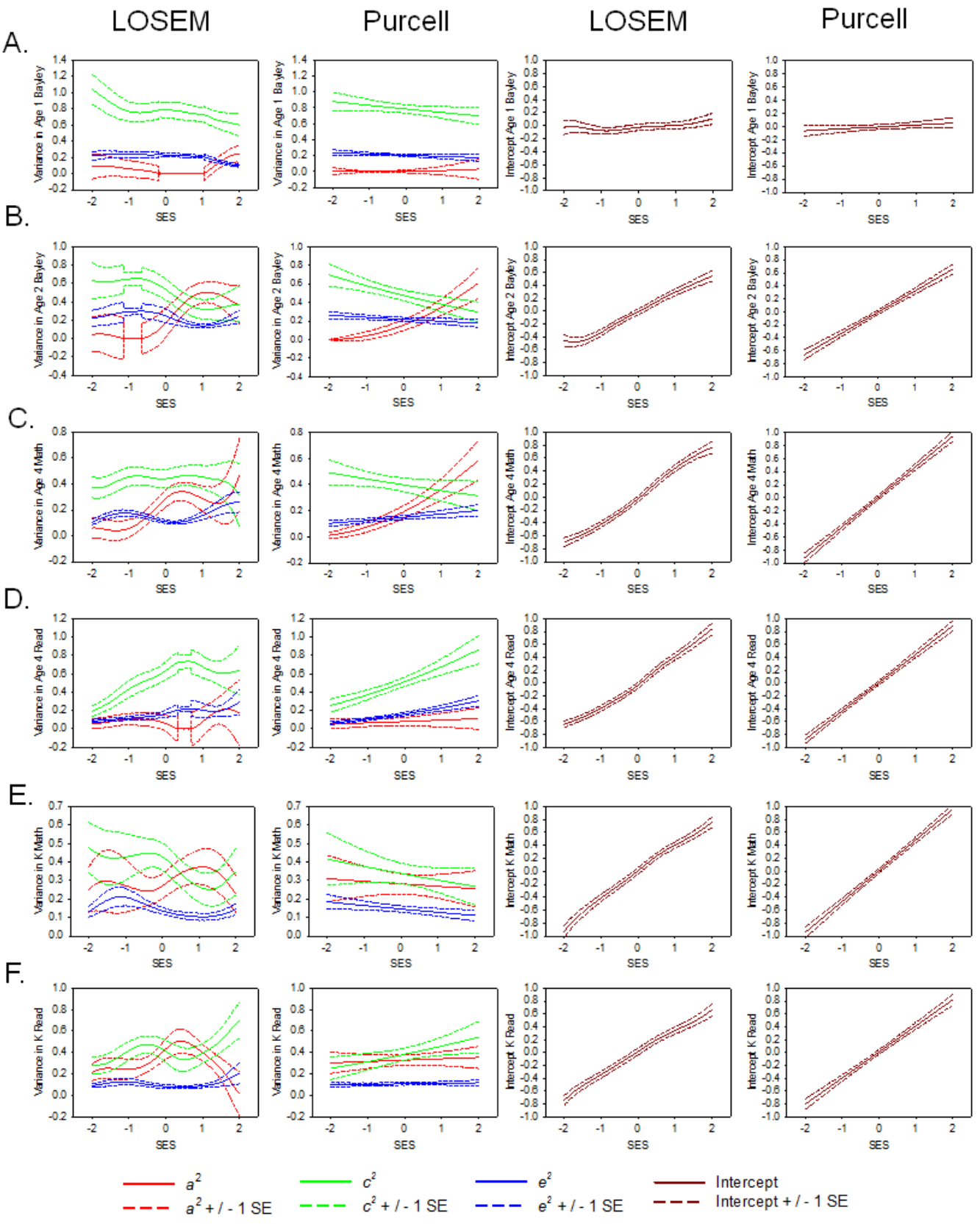
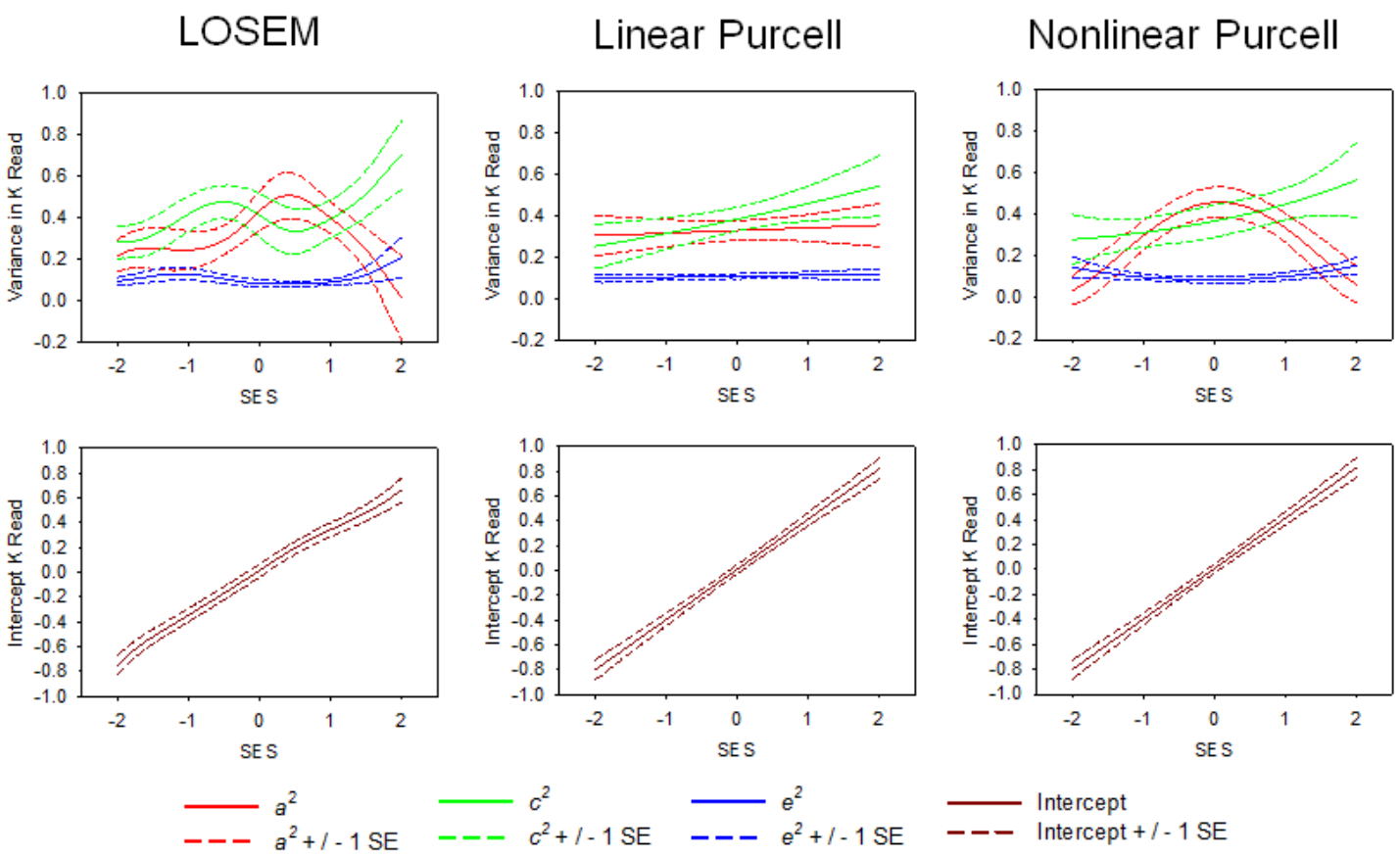

LOSEM
Purcell
LOSEM
Purcell
A.
B.
C.
D.
E.
F.
Variance in Age 1 Bayley
Variance in Age 2 Bayley
Variance in Age 4 Math
Variance in Age 4 Read
Variance in K Math
Variance in K Read
Intercept Age 1 Bayley
Intercept Age 2 Bayley
Intercept Age 4 Math
Intercept Age 4 Read
Intercept K Math
Intercept K Read
SES
a^2
a^2 + / - 1 SE
c^2
c^2 + / - 1 SE
e^2
e^2 + / - 1 SE
Intercept
Intercept + / - 1 SE

Figure 4.


LOSEM
Linear Purcell
Nonlinear Purcell
Variance in K Read
Intercept K Read
SES
a^2
a^2 + / - 1 SE
c^2
c^2 + / - 1 SE
e^2
e^2 + / - 1 SE
Intercept
Intercept + / - 1 SE

Nonparametric Estimates of Gene × Environment Interaction Using Local Structural Equation Modeling

Online Supplement

In this online supplement, we provide additional details about our methodological approach, model fit statistics, parameter estimates, and example scripts to carry out local structural equation modeling (LOSEM).

Methodological Approach

We used data from the Early Childhood Longitudinal Study – Birth Cohort (ECLS-B; Snow et al., 2009). This sample is nationally representative of children born in 2001 in the United States and includes families from a wide range of socioeconomic contexts. The twin subsample includes approximately 650 pairs of twins, of which approximately 150 are monozygotic pairs and 450 are dizygotic pairs.[5] Zygosity was determined by physical similarity ratings made by trained interviewers. For all analyses, we used a socioeconomic status (SES) variable that was based on parental education, occupational prestige, and household income at the baseline wave. We used age 10 months Bayley mental development index, age 2 years Bayley mental development index, age 4 years math readiness, age 4 years reading readiness, kindergarten math achievement, and kindergarten reading achievement as indicators of cognitive ability. All cognitive measures are highly reliable and well-validated (see Snow et al., 2009 for full description). All variables were standardized prior to analysis.

To apply the LOSEM approach, we used a classical univariate twin model. This model decomposes variance in a phenotype into that which is due to additive genetic effects (A), shared environmental effects (C), and nonshared environmental effects (E) by using the known variation in genetic similarity across monozygotic and dizygotic twin pairs. This is accomplished by

[5] Sample sizes are rounded to the nearest 50 in compliance with ECLS-B data restrictions.

specifying that the variance in the observed phenotype is partitioned between three latent ACE factors created for each member of the twin pair. For monozygotic twins, the correlation between A factors is set to 1.0, and for dizygotic twins, the correlation between A factors is set to 0.5. This reflects the amount of shared segregating genetic material between siblings. Shared environmental factors are set to correlate at 1.0, by definition, and nonshared environmental factors are set to correlate at 0.0, by definition. Following this specification, the pathways from the latent ACE factors to the phenotype can be used to infer genetic and environmental influences. This model acts as the baseline model for the LOSEM approach. Then, models are fit to the data at different target levels of the moderator (SES) to produce smoothed estimates of genetic and environment influences.

The Purcell (2002) model builds on this baseline model to explicitly include the moderator (SES) in the model. The main effect of the moderator on the phenotype is controlled. Then, the pathways from the latent ACE factors are constrained to be a linear function of the moderator (e.g., b + b'*SES) which produces quadratic variance components with respect to the moderator. To evaluate nonlinear variance component trends, a quadratic term can be included (e.g., b + b'*SES + b''*SES^2) to produce quartic variance components with respect to the moderator.

We applied these two approaches to test gene × environment interaction to each of the six cognitive phenotypes available in the ECLS-B dataset.

Model Fit Statistics

Supplementary File S1 presents model fit statistics for all LOSEM models. All models fit the data well except for LOSEM models targeted at very high levels of SES for age 4 reading.

Parameter Estimates

Supplementary File S2 presents parameter estimates for all LOSEM models, and Table S1 presents parameter estimates for all Purcell (2002) models.

Example Scripts

Appendix S1 presents an example template file intended to be used with the MplusAutomation package. The template file is very similar to a standard input file, but includes a section specifying certain iterators to create multiple input files. Appendix S2 presents an example *R* script for using the MplusAutomation template file and creating an aggregated datafile. This section is very similar to the example vignettes associated with the MplusAutomation package. We encourage readers to explore this material to see the full functionality of the package. Appendix S3 presents an example *R* script for using OpenMx to perform the LOSEM approach. The analysis is accomplished by using a pre-written function (umxGxE_window) specifically designed to carry out the LOSEM procedure on twin data. All scripts assume a datafile with individual participant families represented as rows, and columns representing the variables for zygo (zygosity; 1=MZ, 2=DZ), Y1 (standardized cognitive score for twin1), Y2 (standardized cognitive score for twin2), and zSES (standardized socioeconomic status).

Table S1. Parameter estimates from Purcell (2002) models

	a	a'	a''	c	c'	c''	e	e'	e''	SES	SES^2
Age 10	.053	.062		.887	-.025		.453	-.018		.030	
months	(.174)	(.142)		(.030)	(.029)		(.017)	(.021)		(.037)	
Bayley				***			***				
Age 2 years	.396	.193		.691	-.072		.468	-.023		.329	
Bayley	(.075)	(.055)		(.042)	(.038)		(.023)	(.020)		(.035)	
	***	***		***			***			***	
Age 4 years	.438	.164		.631	-.035		-.387	-.032		.465	
Math	(.055)	(.043)		(.042)	(.038)		(.021)	(.019)		(.034)	
	***	***		***			***			***	
Age 4 years	.284	.026		.717	.106		.402	.075		.441	
Read	(.081)	(.063)		(.035)	(.034)		(.022)	(.019)		(.033)	
	***			***	**		***	***		***	
K Math	.531	-.013		.582	-.032		.382	-.024		.476	
	(.052)	(.046)		(.048)	(.046)		(.023)	(.019)		(.035)	
	***			***			***			***	
K Read	.573	.011		.619	.059		.328	.008		.406	
	(.042)	(.039)		(.047)	(.046)		(.019)	(.015)		(.037)	
	***			***			***			***	
K Read	.678	.015	-.115	.608	.056	.008	.296	.003	.023	.405	-.004
nonlinear	(.054)	(.045)	(.046)	(.064)	(.036)	(.031)	(.027)	(.018)	(.016)	(.034)	(.029)
	***		*	***			***			***	

Notes. Parameters not labeled with represent the baseline ACE effect. Parameters labeled with ' represent the interaction effect (× SES). Parameters labeled with " represent the quadratic interaction effect (× SES^2). Point estimates are presented with standard errors in parentheses.

* $p < .05$; ** $p < .01$; *** $p < .001$.

Appendix S1 – Template File

```
[[init]]            !This section defines varying iterators.
iterators = mod;    !Here, the iterator is a variable called mod.
mod = 100:500;      !Mod is a vector from 100 to 500.
filename = "[[mod]] SES for Age 1 Bayley.inp"  !Multiple files with unique names.
outputDirectory = "C:/Mplus_Automation/Age 1"; !Target file pathway.
[[/init]]

TITLE: [[mod]] SES for Age 1 Bayley;    ![[mod]] indicates each level of the iterator
                                        !will be used to create multiple input files.

DATA: FILE IS "data.dat"

DEFINE:

!Rescale a standardized SES variable to have -2 to +2 SES equal 100 to 500.
!The init section does not allow for other functions, such as seq().
!Add a positive constant larger than the smallest negative value.
!Multiply by 100 so that each iteration will increment .01 of the original scale.

ses100 = (zSES + 3) * 100;

!Specify the LOSEM weighting approach.

!bandwidth = 2*N^(-1/5)*SDmod
bw = 2*650^(-1/5)*100

!scaled distance = (moderator – target level of moderator)/bandwidth
zx = (ses100 - [[mod]])/bw;
!Note the inclusion of [[mod]] specifies this will vary from 100 to 500.

!kernel weights = (1/(2pi)^.5)*exp(-scaled distance^2/2)
k = (1/(6.283185^.5))*exp((-(zx^2))/2);

!weight = k / .399.
w = k/.399;

VARIABLE:
NAMES ARE
zygo
Y1
Y2
zSES;

MISSING ARE ALL (-9999);
USEVARIABLES ARE Y1 Y2 w;  !Y1 and Y2 are standardized cognitive phenotypes.
WEIGHT = w;                !Weight based on target levels of SES.
GROUPING IS zygo (1=mz 2=dz);

ANALYSIS:
MODEL = NOCOVARIANCES;
TYPE = COMPLEX;

MODEL:
!Standard univariate ACE model.
xA1 by x1*(lxa); xA1@1; [xA1@0];
xA2 by x2*(lxa); xA2@1; [xA2@0];

xC by x1*(lxc); xC@1; [xC@0];
xC by x2*(lxc);
```

```
xE1 by x1*(lxe); xE1@1; [xE1@0];
xE2 by x2*(lxe); xE2@1; [xE2@0];

x1@0;
x2@0;

[x1*](mx);
[x2*](mx);

MODEL MZ:
xA1 WITH xA2@1;

MODEL DZ:
xA1 WITH xA2@.5;
[xA1@0];
[xA2@0];

!The model constraint command can be used to calculate the variance components.
!This also facilitates the extraction of key parameters in the next step.
MODEL CONSTRAINT:
NEW(lla llc lle h2 c2 e2 mean);

lla = lxa;
llc = lxc;
lle = lxe;
h2 = lxa^2;
c2 = lxc^2;
e2 = lxe^2;
mean = mx;

!Save this file as "C:/Mplus_Automation/Age 1 Template File.inp"
```

Appendix S2 – MplusAutomation *R* Script

```
#Load (or install) the MplusAutomation package
library(MplusAutomation)

#Create input files for each level of the moderator
createModels("C:/Mplus_Automation/Age 1 Template File.inp")

#Run all 401 input files. Ensure that the datafile is in the folder with the files.
runModels("C:/Mplus_Automation/Age 1")
#Note, this location is referenced in template file as the outputDirectory for the
#input files.

#Put all model parameters in the age1 object.
age1<-extractModelParameters("C:/Mplus_Automation/Age 1")

#Pull out the unstandardized parameters.
unstdage1<-sapply(age1, "[", "unstandardized")

#Pull out and apply variable names.
oldNamesage1<-names(age1)
names(unstdage1)<-oldNamesage1

#Restructure.
lapply(names(unstdage1), function(element){unstdage1[[element]]$filename<<-element})

#Merge separate models.
mage1<-do.call("rbind",unstdage1)

#Pull out the reduced desired parameters (e.g., the model constraint section).
rage1<-mage1[mage1$paramHeader=="New.Additional.Parameters",]

#Label the parameters with SES levels.
rage1$SES<-rep(100:500, each=7)

#Load (or install) the reshape package.
library(reshape)

#Create long format file.
longa1<- melt(rage1, id.vars = c('SES','param'),measure.vars =
c('est','se','est_se','pval'))

#Create wide format file.
widea1<-cast(longa1,SES~param+variable)

#Pull out model fit statistics.
fita1<- extractModelSummaries("C:/Mplus_Automation/Age 1")

#These last two commands create datafiles that have 401 rows reflecting models for
#each target level of the moderator. The columns reflect the different parameters,
#standard errors, or fit statistics. These objects can easily be used to plot trends
#in R or exported. Reduced versions of these files are presented as Supplementary File
#1 for model fit statistics and Supplementary File 2 for parameter estimates.
```

Appendix S3 – *R* Script for OpenMx

```
#Load OpenMx 2.0
library(OpenMx)
#Install helper library for OpenMx (requires devtools to be installed)
if(library("devtools")){
       install.packages("devtools") #install devtools
       library("devtools")
}
if(library("umx")){
       install_github("tbates/umx") #install umx
       library("umx")
}

#Read data.
data<-read.csv("C:/data.csv", header = TRUE, sep=",",)

#Select dependent variables.
selDVs        = c("Y1", "Y2")

#Select moderator
moderator     = "zSES"

#Exclude participants that are missing on the moderator
data          = data[!is.na(data[,moderator]),]

#Subset MZ and DZ data
mzData        =subset(data, ZYGO =="1",c(selDVs, moderator))
dzData        =subset(data, ZYGO =="2",c(selDVs, moderator))

#Define LOSEM increments
targets       = seq(from = -2, to = 2, by =.01)

#Run and plot for specified windows
umxGxE_window(selDVs = selDVs, moderator = moderator, mzData = mzData, dzData =
dzData, specifiedTargets = targets)
```